**On THE PAPER TITLED “STATISTICAL MODELING: THE 2 CULTURES” BY PROFESSOR LEO BREIMAN**

In this paper, the author examined the problems associated with the over-insistence of core statisticians on data models. According to the paper, 98% of statisticians are in the data modeling culture while only 2% are in the algorithmic model community. The author argued that the focus of the statistical community on data models has led to irrelevant theories, kept statisticians from taking advantage of algorithmic models, and prevented them from working on exciting new projects. The author’s opinion and interest in this discourse emanated from his initial academic experience before he moved into consulting, his experience as a consultant in the industry, and his experience at UC, Berkeley after he returned to the academic community. Based on this movement between academia and industry, the author was able to experience both cultures and form an informed opinion.

The author discussed a few of the projects he executed while in the industry as a consultant and how these projects shaped his cultural beliefs. He emphasized his Ozone Prediction Project and Chlorine Project. The Zone Prediction Project, although failed, the author believes the project will be a success if revisited today considering the now available sophisticated algorithmic models. Following the author’s experience in the industry as a consultant, he went back into the academic community with the perceptions that consultants focus on finding solutions, searching for models that give better solutions, be it algorithmic or data models, and accuracy of prediction of test data as criteria for assessing model. On the other hand, statisticians in the academic community continued to emphasize the goodness of fit and residual assessment of data models as the determinant of model performance.

The author identified some of the common problems with data model which are failure of the goodness of fit test and residual analysis in model assessment. The failure of goodness of fit test is due to the yes/no nature of the test while residual analysis also fail when data dimension is more the 4 to 5. The author agreed that predictive accuracy which is the parameter for measuring the performance of the algorithmic models is most appropriate for measuring the performance of models.

Finally, the author pointed out the increasing popularity and applicability of algorithmic model especially with the birth of neural networks and decision tree. Emphases of these algorithmic models are in achieving predictive accuracy and this accuracy is measured by cross-validation which the author believe is more effective than its equivalent in the data modeling community. The popularity of algorithmic models has continued to grow with the recent emergence of ensemble techniques like random forest which involve the combination of the predictive capability of multiple models to improve model performance. The author wants statisticians to reconsider their preference for data models.